
Chapter 2: Sample Design & Fielding Procedures

2.1 Sample Design

Because this section contains a large amount of detail, a rough overview of the sample design is first provided. In 1978, the National Opinion Research Center (NORC) at the University of Chicago created a list of housing units in selected areas of the United States for the first NLSY79 interview. NORC interviewers went to a random sample of these homes and performed a short interview called the screener, which provided basic information on every resident of the household. NORC also extracted a random sample of Department of Defense records to ensure the survey covered members of the military.

Together these two processes provided information, such as name, age, sex, race, and address, on more than 155,000 people. This information was then used to identify all individuals aged 14 to 21 as of December 31, 1978. Based on this screener information, each appropriately aged individual was assigned to one of the sample groups. Then, in 1979, NORC interviewers asked individuals on this list to participate in the first NLSY79 interview. Any person who completed the first round interview is considered a member of the NLSY79 cohort.

Three independent probability samples comprise the NLSY79. These samples are designed to represent the entire population of youth aged 14 to 21 as of December 31, 1978, residing in the United States on January 1, 1979. The three samples are:

- (1) a cross-sectional sample (6,111) designed to be representative of the noninstitutionalized civilian segment of young people living in the United States in 1979 and born January 1, 1957, through December 31, 1964
- (2) a set of supplemental samples (5,295) designed to oversample civilian Hispanic, black, and economically disadvantaged, non-Hispanic, non-black youth born in the same time period
- (3) a military sample (1,280) designed to represent the population born January 1, 1957, through December 31, 1961, serving in the military as of September 30, 1978. The inclusion of the military sample allows comparative civilian/military analyses by ensuring more than the pro rata share of cohort members in the military

Users can identify a respondent's sample type by using variable R01736. Beginning in 1986, additional information was collected about children born to female NLSY79 respondents. The child sample, when weighted, is representative of American children born to the population of women born in 1957 through 1964 and living in the United States in 1979. The sampling procedures used by NORC to select the civilian and military subsamples differed and are discussed separately below. For additional information on NLSY79 sampling procedures, see Frankel et al. (1983) and the *Interviewer's Manual* for the 1978 household screening (NORC 1978). Sampling issues related to the Children of the NLSY79 are discussed in Baker et al. (1993) and in the *NLSY79 Child & Young Adult Data Users Guide*.

2.2 Screening

To find people of the correct age and ethnic composition, NORC screened a large number of individuals in two separate procedures. First, household screening interviews were conducted to select the NLSY79 civilian cross-sectional and supplemental subsamples from the civilian population. Then, a second screening was done to identify the military sample. While the civilian screening made use of field interviewers going to preselected households, the military sample was drawn from Department of Defense internal records.

NORC administered the civilian sample screening interview in approximately 75,000 dwellings and group quarters. These interviews occurred in 1,818 sample segments of 202 Primary Sampling Units (PSUs), which included most of the fifty states and the District of Columbia. The screening interview was designed to elicit information that would identify persons eligible for inclusion in the NLSY79 sample. The civilian screening interviews were completed within 91.2 percent of the cross-sectional and 91.9 percent of the supplemental occupied dwelling units selected for screening.

Cross-Sectional Sample: Approximately 18,000 of the screening interviews were carried out among 918 sample segments in 102 Primary Sampling Units (PSUs), which were selected from the NORC Master Probability Sample of the United States.

Supplemental Sample: A total of 57,000 screening interviews for the supplemental sample were carried out among 900 sample segments in a 100-PSU sample specifically designed to produce statistically efficient samples of Hispanics, blacks, and economically disadvantaged, non-black/non-Hispanics.

The NLS sample design, which selected every eligible person connected to the household, generated a representative sample of siblings and spouses living in the same household and satisfying the age restrictions stated above. However, NLSY79 samples do not contain nationally representative samples of siblings and spouses of all ages and living arrangements. When the NLSY79 is used to study sibling pairs and married couples, care must be used in generalizing from the findings of such studies.

Procedures were also developed to establish “linkages” between dwellings and certain types of individuals who might be temporarily absent. As part of the initial screening for the civilian sample, household respondents were asked if there were any persons with primary family connections to the household who were away from the household at the time. Included in this group were college students, military personnel, and those in prisons or other institutions. Household screener respondents were also asked to name persons who might occasionally stay at the dwelling who did not have any other “usual place of residence.” For each individual identified in this process, an attempt was made to determine whether the individual would be “linked” to some other household, e.g., college students living off campus in their

own dwelling units. All individuals without other linkages were included in the household composition for purposes of subsampling.

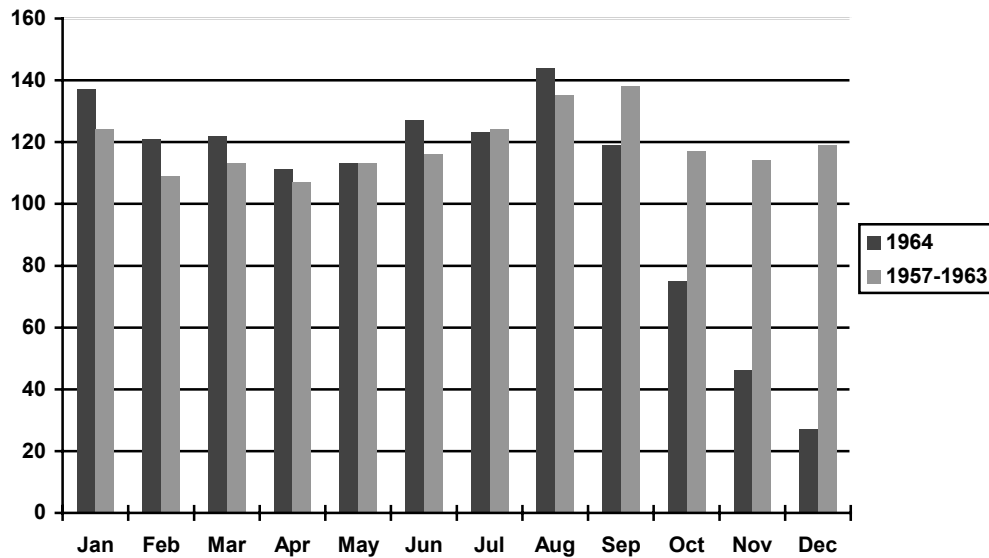
Military Sample: Persons on active military duty as of September 30, 1978, were sampled from rosters provided by the Department of Defense. No formal screening interview was conducted.

2.3 Sampling Process

Civilian Samples: All civilian sample selection was accomplished through a multi-stage stratified area probability sample of dwelling units and group quarter units. A moderate degree of oversampling of dwelling units within sample listing segments was employed in order to increase the sample composition with respect to the targeted groups of the supplemental sample.

Base year samples of Hispanics, blacks, and economically disadvantaged, non-blacks, non-Hispanics were selected from individuals identified in both the 102 PSU cross-sectional sample and the 100 PSU special purpose sample. To the extent that individuals identified in the screening phase were obtained with different probabilities of selection (because of selective oversampling), the weighting of base year samples attempts to minimize these probability differences. Since the use of oversampling tends to decrease sample efficiency (i.e. variance), attempts were made to hold required oversampling to a minimum.

User Notes: At all selected dwellings, attempts were made to obtain appropriate classification information for all persons living in the dwelling. In order to minimize the potential for “interviewer effect,” survey interviewers were not informed about specific groups that would be included in the subsequent interviews. However, the distribution of month of birth by birth year departs from randomness for the youngest members of the cohort, those born in 1964 (refer to Figure 2.3.1). This non-randomness most likely comes from two sources. First, some of the screening was done by supervisors and other higher level staff who were familiar with the specific age groups that belonged in the survey; this could have introduced interviewer bias. Second, families who wanted to find out more information could contact NORC or the Department of Labor and find out the age ranges the survey was trying to capture. This extra information could have led to non-random self-selection at the edge of the age range.

Figure 2.3.1 Number of Respondents Born Each Month by Birth Year ¹

¹ The month and year of birth is taken from the 1978 screener (R00003. and R00005.). Respondents were asked about their dates of birth again in 1981, but the use of these values does not change the results indicated above. The 1957–63 value can be found by averaging the total number of birth dates reported for each month over all of the years.

Assignment of a youth to either the cross-sectional or supplemental sample, was made using information collected during the household screening interviews and a set of coding instructions prepared by NORC. During the 1978 household screening interviews—from which the sample of NLSY79 respondents was drawn—information was collected on the sex, race, and ethnic origin of each household member and on the total income of the family unit during the past 12 months. A detailed set of coding procedures transformed these raw data into a combined racial/ethnic identifier and an economically disadvantaged qualifier. These criteria were used not only to assign a youth to one of the various subsamples, e.g., the economically disadvantaged, non-black/non-Hispanic supplemental sample, but also to specify the primary race/ethnicity variable, which provides the basis for weighting.

Other technical information on the sample assignment process can be found in (1) the *Household Screener and Interviewer's Manual* (NORC 1978), which includes a copy of the screening instrument and detailed instructions to interviewers for administering the race, ethnic origin, and family income questions; (2) the *Technical Sampling Report* (Frankel et al. 1983), which describes the NLSY79 sample selection procedures for the civilian and military subsamples; (3) the 10/4/78 NORC memorandum, which provides the rules used to assign race and poverty status from responses to the screening questions; (4) a copy of the 1978 poverty income levels by family size and farm-nonfarm residence; and (5) the “Race, Ethnicity, & Nationality” section of this *User's Guide*, which summarizes information in these documents as it relates to the assignment of “Hispanic,” “Black,” and “Non-Hispanic, Non-Black” origins used in the sample

identification code variable (R01736.) and the race/ethnicity variable (R02147.). Base year interviews with the three subsamples were conducted between January and mid-August 1979. Table 2.3.1 summarizes base year completion rates for each subsample.

Table 2.3.1 Base Year Interview Completion: NLSY79

	Designated for Interviewing	Interviewed Initial Survey Year	
Total Cohort	14574	12686	87%
Cross-Sectional Sample ¹	6812	6111	90%
Supplemental Sample ¹	5969	5295	89%
Military Sample	1793	1280	72%

¹ As determined through the household screening.

Cross-Sectional Sample: Following the initial screening process, 6,812 individuals from the cross-sectional sample were designated to be interviewed in the base year; of those, 90 percent or 6,111 respondents were actually interviewed in 1979. The cross-sectional sample is designed to maximize the statistical efficiency of samples which are “cross-sectional” with respect to the rural population. Specifically, through the several stages of sample selection (counties, enumeration districts-block groups, sample listing units), probabilities of selection are based upon either total population or total housing units. Except for the economically disadvantaged supplemental sample, sampling of non-black/non-Hispanic respondents was restricted to the 102 PSU National Sample.

Supplemental Sample: After screening, 5,969 individuals from the supplemental sample were designated for base year interviews, and of these, 89 percent or 5,295 respondents were actually interviewed. Stratification specifically relevant for Hispanics, non-Hispanic blacks, and economically disadvantaged, non-black/non-Hispanics was used. Probability proportional to size procedures were based on size measures for these groups rather than for the general population, making it possible to more nearly equalize the distribution of the targeted groups among the various sampling units than would otherwise be the case.

Military Sample: Of the 1,793 military youth selected for interviews, 1,280 or 72 percent were interviewed in 1979. Selection of the military sample was accomplished in two stages. In the first stage, a sample of approximately 200 “military units” was selected. These units were selected with probabilities proportional to the number of persons born in 1957 through 1961 and serving in the military unit as of September 30, 1978.

Within selected units, persons born in 1957 through 1961 were subsampled with probabilities inversely proportional to the first-stage selection probability. Females were oversampled at a rate approximately six

times that of males in order to produce approximately 850 males and 450 females. Within each sex, the sample was stratified on the basis of branch of military service (Army, Navy, Air Force, and Marine Corps) and geographic location (Eastern U.S., Western U.S., Europe, Far East, other). Of those interviewed in 1979, 824 military respondents were male and 456 were female (see Table 2.3.2). The entire military sample was eligible for interview from 1979–84.

**Table 2.3.2 NLSY79 Military Respondents Interviewed in 1979
by Gender, Race & Military Branch**

	Total	Males	Females
Total Military	1280	824	456
Non-black/non-Hispanic	951	609	342
Black	251	162	89
Hispanic	78	53	25
Military Branch			
Army	578	354	224
Navy	280	212	68
Air Force	293	162	131
Marine Corps	129	96	33

Child Sample: The number of children assessed during a given child survey year is a function of the number of children born to interviewed NLSY79 mothers, the number of children living in the homes of those mothers, and, finally, the number of those children actually interviewed. Of the 5,842 NLSY79 females eligible for the first child interview in 1986, more than 2,900 mothers and 4,971 children were interviewed. From this sample of eligible children, assessment data were collected for 4,786. Table 1.6.1 in Chapter 1 presents sample sizes across subsequent survey years.

2.4 Interview Schedule & Fielding Periods

The original interview schedule, which called for yearly personal interviews with NLSY79 respondents, was retained from 1979 through 1986. In 1987, budget constraints dictated a limited phone interview rather than a personal interview. Personal interviews resumed with the 1988 round and continued yearly until 1994. Since 1994, NLSY79 respondents have been interviewed every other year (1996, 1998, etc.).

The initial NLSY79 interviews were conducted between late January and mid-August 1979. The next several interviews were fielded in the first six months of the year; subsequent surveys have typically begun in late spring and ended in the fall or early winter. Table 2.4.1 provides information on the fielding periods for the youth and child samples.

Table 2.4.1 Fielding Periods: NLSY79 & NLSY79 Children

Cohort	Survey Year(s)	Fielding Period
NLSY79	1979–80	January–August
	1981–82	January–July
	1983–85	January–June
	1986	February–July
	1987	March–October
	1988–91	June–December
	1992	May–December
	1993	June–November
	1994	June–December
	1996	April–October
	1998	March–September
	2000	April 2000–January 2001
NLSY79 Children	1986	February–July
	1988	June–December
	1990	July–December
	1992	May–December
	1994	June–December
	1996	April–October
	1998	March–September
	2000	May 2000–January 2001

From 1979 until 1986, timing of the fielding period was designed to allow all respondents still in school to be interviewed before they left to take temporary summer jobs. Detailed information was collected for jobs held by respondents while they were in school. Since the youngest respondents in the survey were 23 years old in 1988, the shift in fielding periods after 1987 had a relatively small impact on information on jobs held while in school. An attempt was made during the initial survey years to keep the fielding period for an individual respondent approximately the same from year to year in order to assure that the time between interviews was approximately twelve months.

Researchers conducting analyses on topics where time periods are critical should carefully examine the reference period of the questions, the actual interview date, and the duration since the preceding interview.

2.5 Interview Methods & Target Universe

During each survey round, NORC attempts to reach all youth within the active samples. No respondents have been routinely excluded from locating efforts with the exception of respondents who have died or, in a small number of cases, were judged to be extremely difficult. The permanent NLSY79 sample

designated for interviewing during the 1979–84 interview years consisted of all civilian and military youth who were interviewed in the base year and who were alive at the survey date.

In 1985, when interviewing of the full military sample ceased, the total NLSY79 sample size dropped from 12,686 to 11,607. Retained for interviewing in post-1984 surveys were 201 military respondents randomly selected from the entire military sample of 1,280; the remaining 1,079 military respondents were eliminated from the sample. The 201 military members who were retained included (1) 51 cases that would have been selected as part of a random sample of youth including the military and (2) 150 additional cases selected to provide a sufficient number of original military sample members to avoid overly large sampling variability for the military sample. Beginning in 1991, the 1,643 members of the economically disadvantaged, non-black/non-Hispanic supplemental sample were no longer interviewed. Unless otherwise noted, eligible sample sizes reported in NLS publications include deceased and difficult-to-field respondents but exclude those respondents dropped from the sample. Additional information on numbers and characteristics of noninterviewed respondents can be found in the “Reasons for Noninterview” section later in this chapter.

NLSY79 respondents reside in each of the 50 states as well as the District of Columbia, U.S. territories, and countries abroad. Locating respondents is a coordinated effort of NORC’s central office, its locating shop, and local-level field staff. Prior to fielding, NORC’s central office sends a short, informative “advance letter” to each respondent reminding him/her of the upcoming interview and confirming the respondent’s current address and phone number. Field staff locating efforts begin with this information and locator sheets.

At the local level, interviewers are responsible for contacting all respondents in their caseloads and for tapping additional local resources (post offices, departments of motor vehicles and vital statistics, etc.) to locate those respondents who have moved. If an interviewer is unsuccessful in locating a respondent, the case is transferred to the field manager who undertakes additional locating strategies.

In the event that such local-level efforts fail, the case is forwarded to NORC’s locating shop in Chicago where the complete hard copy files on each respondent can be accessed and used for additional locating efforts. Respondents who cannot be located are only a small percentage of the total not interviewed in a given survey year. (For more information about noninterviews, refer to section 2.6 in this chapter.)

In addition to its comprehensive locating efforts, NORC makes every effort to convert initial respondent refusals to completed interviews. For uncooperative respondents, NORC sends “refusal conversion letters” designed to encourage continued participation in the survey. These letters are often written by field managers who have personal knowledge of specific respondents and can customize the letter to an

individual respondent's concerns. Over the years, NORC has successfully conducted interviews with 33–50 percent of respondents who initially refused. NORC's extensive locating methods and its conversion strategy, combined with its close monitoring of response rates for each of the subsamples of the NLSY79, have resulted in relatively high retention rates for a longitudinal panel of this duration.

While personal interviewing has remained the primary contact method used for all but one of the NLSY79 surveys, it is not the exclusive method. Telephone contact within personal survey rounds occurs under certain circumstances, e.g., where the respondent resides in a remote area or field staff determines that phone contact is the preferred method of interviewing a specific respondent. The number of telephone interviews has increased in recent years as respondents have become more dispersed or expressed a preference for phone interviews. The percent of surveys conducted by telephone for each survey year are shown in table 2.5.1.

Table 2.5.1 Percent of NLSY79 Interviews Conducted by Telephone, 1979–2000

Year	Number of Phone Interviews	Percent of Total Interviews	Year	Number of Phone Interviews	Percent of Total Interviews
1979	548	4.4	1989	1518	14.3
1980	648	5.3	1990	1317	12.6
1981	654	5.4	1991	1241	13.8
1982	1054	8.7	1992	1164	12.9
1983	324	2.6	1993 ¹	—	—
1984	646	5.3	1994 ¹	—	—
1985	953	8.7	1996	1042	12.1
1986	929	8.7	1998	2069	24.6
1987	8998	85.8	2000	2613	32.5
1988	920	8.8			

¹ Questions identifying whether interview was conducted by telephone were not included in the 1993 and 1994 surveys.

In rare cases, interviews are conducted in whole or in part with a proxy, a person other than the respondent (for example, four in 1991, two in 1992). A variable, entitled 'Interview Conducted with Proxy Respondent,' is present in the data to identify these interviews. In order to conduct such an interview, individual approval must be obtained by the NORC central office and the circumstances documented.

A Spanish version of all survey instruments, except the Young Adult questionnaire, is prepared and NORC employs bilingual, Spanish-speaking interviewers. During the 1992 interview, for example, 104 respondents requested the use of a Spanish version of the questionnaire.

The average length of a personal interview is approximately one hour. The 1987 telephone interviews were completed within about 40 minutes, while the administration of the child assessments adds approximately 45 minutes to the total survey administration time for each child.

Until 1989, the NLSY79 was conducted using only paper-and-pencil interviews (PAPI). PAPI interviews were performed by interviewers filling in the relevant fields of large printed questionnaire booklets. While these booklets were cheap to produce, interviewers could make mistakes in following complicated skip patterns and filling in answers. Moreover, after all interviews were completed, additional office staff were needed to transcribe the information collected. Computer-assisted personal interviews (CAPI) were designed to eliminate many of these problems. For CAPI interviews, interviewers take laptop computers into the field instead of questionnaire booklets. A computer program automatically selects the next question, prevents interviewers from entering illegal values, and warns interviewers about questionable answers. The computer also eliminates the need for data transcription except for specific items collected verbally and coded later.

While the majority of interviews in 1990 were collected using PAPI materials, a subset of one fourth of respondents were administered each of the surveys using CAPI methods in order to test the viability and reliability of CAPI administration. Due to the success of these experiments, the NLSY79 interviews became fully CAPI administered beginning in 1993. Users interested in the results of these experiments should consult Olsen (1991).

From 1979 until 1994, each respondent was paid \$10 upon completion of the interview. Respondents were paid \$20 beginning with the 1996 interview. NLSY79 mothers participating in the child assessments receive an additional \$5 per child. Through 1994, young adults were paid \$10. In 1996, 1998 and 2000, they received \$20. NLSY79 respondents who participated in the 1980 *ASVAB* testing were paid \$50 each.

Retention vs. Response Rates

Retention rates for NLSY79 respondents throughout the first 16 rounds of the survey exceeded or were very near 90 percent. Rounds 17 and 18 saw retention rates of approximately 85 percent. Retention rate is calculated by dividing the number of respondents interviewed by the number of respondents remaining eligible for interview. All 1979 (round 1) respondents including those reported as deceased are eligible for interviews, with the exception of those who have been permanently dropped from the sample. In the round 19 (2000) survey, 8,033 civilian and military respondents out of the 9,964 eligible were interviewed, for an overall retention rate of 80.6 percent. Retention rates for each survey are shown in table 2.5.2. This table also shows which interviews were conducted with paper-and-pencil interviewing (PAPI) and which were executed with computer-assisted personal interviewing (CAPI).

The number of respondents can also be expressed as a percentage of the number of base year respondents **not** known to be deceased. This is referred to as the response rate and is reported in table 2.5.3 for each survey round. As of 2000, 335 main respondents (more than 3 percent of the respondents eligible for interview) had been reported as deceased. The response rate for those believed to be alive is 83.2 percent.

Table 2.5.2 Sample Sizes & Retention Rates by Sample Type: NLSY79

Year	Type & Mode of Interview	Cross-Sectional Sample		Supplemental Sample		Military Sample		Total Sample	
		Total	Retention Rate ¹	Total	Retention Rate ¹	Total	Retention Rate ¹	Total	Retention Rate ¹
1979	Personal/PAPI	6111	—	5295	—	1280	—	12686	—
1980	Personal/PAPI	5873	96.1	5075	95.9	1193	93.2	12141	95.7
1981	Personal/PAPI	5892	96.4	5108	96.5	1195	93.4	12195	96.1
1982	Personal/PAPI	5876	96.2	5036	95.1	1211	94.6	12123	95.6
1983	Personal/PAPI	5902	96.6	5093	96.2	1226	95.8	12221	96.3
1984	Personal/PAPI	5814	95.1	5040	95.2	1215	94.9	12069	95.1
1985	Personal/PAPI	5751	94.1	4957	93.6	186 ²	92.5	10894 ³	93.9
1986	Personal/PAPI	5633	92.2	4839	91.4	183	91.1	10655	91.8
1987	Telephone/PAPI	5538	90.6	4768	90.1	179	89.1	10485	90.3
1988	Personal/PAPI	5513	90.2	4777	90.2	175	87.1	10465	90.2
1989	Personal/PAPI/CAPI	5571	91.2	4853	91.7	181	90.0	10605	91.4
1990	Personal/PAPI/CAPI	5498	90.0	4755	89.8	183	91.0	10436	89.9
1991	Personal/PAPI	5556	90.9	3281 ⁴	89.9	181	90.0	9018 ⁵	90.5
1992	Personal/PAPI	5553	90.9	3280	89.8	183	91.0	9016	90.5
1993	Personal/CAPI	5537	90.6	3293	90.2	181	90.0	9011	90.4
1994	Personal/CAPI	5457	89.3	3256	89.2	178	88.6	8891	89.2
1996	Personal/CAPI	5290	86.6	3171	86.8	175	87.1	8636	86.7
1998	Personal/CAPI	5159	84.4	3065	83.9	175	87.1	8399	84.3
2000	Personal/CAPI	4949	81.0	2921	80.0	163	81.1	8033	80.6

¹ Retention rate is defined as the percentage of base year respondents within each sample type remaining eligible who were interviewed in a given survey year. Included in the eligible sample are deceased and difficult to field respondents whom NORC does not attempt to contact.

² A total of 201 military respondents were retained from the original sample of 1,280.

³ The total number of civilian and military respondents in the NLSY79 at the beginning of the 1985 survey was 11,607.

⁴ Economically disadvantaged, non-black/non-Hispanic female and male members of the supplemental subsample are not eligible for interview as of the 1991 survey year. Remaining eligible for interview in post-1990 surveys are 3,652 black and Hispanic respondents of the supplemental sample, of whom 3,281 were interviewed in 1991.

⁵ The total number of civilian and military respondents in the NLSY79 at the beginning of the 1991 survey was 9,964.

Table 2.5.3 Response Rates (Excluding Deceased Respondents) by Sample Type: NLSY79

Year	Cross-Sectional Sample			Supplemental Sample			Military Sample			Total Sample		
	#	Response		#	Response		#	Response		#	Response	
	Total	Deceased	Rate	Total	Deceased	Rate	Total	Deceased	Rate	Total	Deceased	Rate
1979	6,111	0	—	5,295	0	—	1,280	0	—	12,686	0	—
1980	5,873	4	96.2	5,075	5	95.9	1,193	0	93.2	12,141	9	95.8
1981	5,892	15	96.7	5,108	14	96.7	1,195	0	93.4	12,195	29	96.3
1982	5,876	24	96.5	5,036	19	95.5	1,211	1	94.7	12,123	44	95.9
1983	5,902	27	97.0	5,093	26	96.7	1,226	4	96.1	12,221	57	96.8
1984	5,814	30	95.6	5,040	33	95.8	1,215	4	95.2	12,069	67	95.6
1985	5,751	36	94.7	4,957	43	94.4	1,186	0	92.5	11,894	79	94.5
1986	5,633	43	92.8	4,839	51	92.3	183	1	91.5	10,655	95	92.6
1987	5,538	51	91.4	4,768	56	91.0	179	3	90.4	10,485	110	91.2
1988	5,513	56	91.0	4,777	68	91.4	175	3	88.4	10,465	127	91.2
1989	5,571	60	92.1	4,853	78	93.0	181	3	91.4	10,605	141	92.5
1990	5,498	67	91.0	4,755	82	91.2	183	3	92.4	10,436	152	91.1
1991	5,556	75	92.0	23,281	87	92.0	181	4	91.9	29,018	166	92.0
1992	5,553	80	92.1	3,280	93	92.1	182	4	92.4	9,015	177	92.1
1993	5,537	89	92.0	3,293	104	92.8	181	4	91.9	9,011	197	92.3
1994	5,457	103	90.8	3,256	117	92.1	178	4	90.4	8,891	224	91.3
1996	5,290	128	88.4	3,171	130	90.0	175	5	89.3	8,636	263	89.0
1998	5,159	151	86.6	3,065	139	87.2	175	5	89.3	8,399	295	86.9
2000	4,949	170	83.3	2,921	158	83.6	163	7	84.0	8,033	335	83.4

Note: Response rate is defined as the percentage of base-year respondents remaining eligible and not known to be deceased who were interviewed in a given survey year.

¹ A total of 201 military respondents were retained from the original sample of 1,280; 186 of the 201 participated in the 1985 interview. The total number of NLSY79 civilian and military respondents eligible for interview (including deceased respondents) beginning in 1985 was 11,607.

² The 1,643 economically disadvantaged nonblack/non-Hispanic male and female members of the supplemental subsample were not eligible for interview as of the 1991 survey year. The total number of NLSY79 civilian and military respondents eligible for interview (including deceased respondents) beginning in 1991 was 9,964.

2.6 Reasons for Noninterview

A ‘Reason for Noninterview’ variable is constructed for each survey year (1980–2000) in the NLSY79 and provides an explanation of why an interview could not be conducted or completed with a respondent. The cause of noninterview is assigned by the NORC interviewer to each respondent designated as a member of the eligible sample for a given survey year. Typical coding categories have included reasons such as an interview being refused by the respondent or by the respondent’s parent, the youth and/or the family unit not being successfully located, or the respondent being reported as deceased.

Beginning in the 1980s, two administrative categories were added. One reflected a decision by NORC not to refield certain cases that were determined to be extremely difficult to interview. The second category indicates that, due to funding cutbacks, interviews would not be attempted with certain members of one or more of the NLSY79 subsamples. Thus, beginning in 1985, interviews ceased for 1,079

respondent members of the military subsample; each was permanently assigned a reason for noninterview of “military sample dropped.” A second group of respondents, those belonging to the supplemental economically disadvantaged, non-black/non-Hispanic sample, was similarly dropped from interviewing beginning with the 1991 survey. The target universe for each survey year—that is, the respondents whom NORC attempts to interview—thus includes all respondents interviewed in the initial survey year exclusive of those who were (1) reported deceased at an earlier interview; (2) dropped from the sample; or (3) judged to be extremely difficult to interview.

User Notes: Reasons for noninterview may change for a given respondent between noninterview years, even if those years are contiguous. Some codes, such as “parent refusal/break off,” have become virtually obsolete over time with the aging of the cohort. Other codes should be considered relatively permanent, such as those applied to the reported death of a respondent. (Users should be aware that false reports of death have been used to avoid being interviewed. NORC attempts to verify these reports by obtaining death certificate information or newspaper obituaries.)

The coding of deceased members of the two subsamples dropped from interviewing in 1985 and 1991 has not been handled consistently. Those respondents of the military sample reported deceased during the 1980–84 surveys, i.e., those with a code of “65 - Deceased” on a ‘Reason for Noninterview’ variable, have been recoded, beginning in 1985, to “68 - Military Sample Dropped”; this recode occurred for four cases. Thus the count of 1,079 reflects all members of the military subsample, both living and deceased, who were dropped from interviewing; however, this means that the cumulative count of total deceased respondents on any post-1984 ‘Reason for Noninterview’ will be understated.

The 22 members of the supplemental economically disadvantaged, non-black/non-Hispanic sample who had died prior to the dropping of the sample in 1991 were not similarly reclassified as dropped. The count of 1,621 for the economically disadvantaged, non-black/non-Hispanic sample in the 1991 ‘Reason for Noninterview’ variable reflects only the living members of the total 1,643 who were dropped; the 22 deceased members of the supplemental economically disadvantaged, non-black/non-Hispanic subsample remain coded as deceased.

The tables below present the number of respondents not interviewed across survey years by reason, gender, race/ethnicity, and sample type.

Table 2.6.1 Reasons for Noninterview: NLSY79 1979–2000

Survey Year	Total Interviewed	Total Not Interviewed	Reason for Noninterview					
			Refusal	Can't Locate	Deceased	Other	Difficult Cases	Dropped ¹
1979	12686	0	0	0	0	0	0	0
1980	12141	545	253	217	9	66	0	0
1981	12195	491	220	114	29	128	0	0
1982	12123	563	177	209	44	50	83	0
1983	12221	465	220	124	57	37	27	0
1984	12069	617	374	151	67	25	0	0
1985	10894	1792	331	152	79	26	125	1079
1986	10655	2031	524	200	95	36	97	1079
1987	10485	2201	508	293	110	68	143	1079
1988	10465	2221	587	248	127	83	97	1079
1989	10605	2081	525	188	141	46	102	1079
1990	10436	2250	662	246	152	52	59	1079
1991	9018	3668	507	202	165	23	71	2700 ²
1992	9016	3670	526	158	177	29	80	2700
1993	9011	3675	541	122	197	24	91	2700
1994	8891	3795	604	146	224	26	95	2700
1996	8636	4050	708	184	263	48	147	2700
1998	8399	4287	731	272	295	136	153	2700
2000	8033	4653	1037	353	335	123	105	2700

¹ Two groups of NLSY79 respondents have been dropped from interviewing: (1) 1,079 members of the 1,280 military subsample were dropped after the 1984 survey and (2) the 1,643 members of the supplemental economically disadvantaged, non-black/non-Hispanic subsample were dropped after the 1990 interview.

² The total number of respondents dropped as of the 1991 survey year is actually 2,722. Twenty-two respondents of the 1,643 supplemental economically disadvantaged, non-black/non-Hispanic sample members dropped after 1990 had been reported as deceased prior to 1991 and remain so classified.

Table 2.6.2 Reasons for Noninterview by Gender: NLSY79 1980–2000

Survey Year	Total Not Interviewed		Reason for Noninterview											
			Refusal		Can't Locate		Deceased		Other		Difficult Case		Dropped ¹	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F
1980	311	234	134	119	129	88	4	5	44	22	0	0	0	0
1981	272	219	109	111	61	53	18	11	84	44	0	0	0	0
1982	315	248	103	74	116	93	30	14	35	15	31	52	0	0
1983	255	210	110	110	71	53	36	21	26	11	12	15	0	0
1984	348	269	186	188	96	55	44	23	22	3	0	0	0	0
1985	1032	760	172	159	93	59	52	27	20	6	57	68	638	441
1986	1166	865	271	253	122	78	65	30	26	10	44	53	638	441
1987	1287	914	273	235	186	107	77	33	43	25	70	73	638	441
1988	1250	971	296	291	138	110	89	38	43	40	46	51	638	441
1989	1207	874	268	257	123	65	96	45	31	15	51	51	638	441
1990	1291	959	322	340	160	86	103	49	34	18	34	25	638	441
1991	1932	1736	264	243	136	66	114	51	16	7	33	38	1369	1331
1992	1922	1748	266	260	106	52	122	55	19	10	40	40	1369	1331
1993	1939	1736	276	265	86	36	140	57	19	5	49	42	1369	1331
1994	1992	1803	310	294	99	47	156	68	13	13	45	50	1369	1331
1996	2128	1922	352	356	124	60	185	78	28	20	70	77	1369	1331
1998	2303	1984	377	354	189	83	207	88	85	51	76	77	1369	1331
2000	2483	2170	513	524	238	115	230	105	75	48	58	47	1369	1331

¹ Two groups of NLSY79 respondents have been dropped from interviewing: (1) 1,079 members of the 1,280 military subsample were dropped after the 1984 survey and (2) the 1,643 members of the supplemental economically disadvantaged, non-black/non-Hispanic subsample were dropped after the 1990 interview. Of the 1,643 supplemental economically disadvantaged, non-black/non-Hispanic subsample members dropped, 22 remain classified as deceased.

Table 2.6.3 Reasons for Noninterview by Race/Ethnicity: NLSY79 1980–2000

Key H: Hispanics, B: Blacks, NB/NH: Non-Black/Non-Hispanics

Survey Year	Total Not Interviewed			Reason for Noninterview																	
				Refusal			Can't Locate			Deceased			Other			Difficult Cases			Dropped¹		
	H	B	NB/ NH	H	B	NB/ NH	H	B	NB/ NH	H	B	NB/ NH	H	B	NB/ NH	H	B	NB/ NH	H	B	NB/ NH
1980	112	124	309	33	50	170	65	57	95	4	2	3	10	15	41	0	0	0	0	0	0
1981	93	94	304	24	42	154	36	22	56	7	8	14	26	22	80	0	0	0	0	0	0
1982	120	120	323	35	35	107	57	54	98	9	11	24	15	7	28	4	13	66	0	0	0
1983	95	110	260	37	50	133	40	31	53	10	15	32	7	10	20	1	4	22	0	0	0
1984	105	126	386	46	64	264	40	37	74	13	16	38	6	9	10	0	0	0	0	0	0
1985	181	331	1280	60	58	213	40	50	62	14	22	43	8	6	12	18	22	85	41	173	865
1986	237	384	1410	89	98	337	62	57	81	16	30	49	11	11	14	18	15	64	41	173	865
1987	292	424	1485	101	92	315	82	89	122	22	35	53	21	15	32	25	20	98	41	173	865
1988	304	432	1485	115	118	354	81	74	93	23	42	62	26	18	39	18	7	72	41	173	865
1989	235	404	1442	91	99	335	56	51	81	25	48	68	6	19	21	16	14	72	41	173	865
1990	273	455	1522	110	133	419	71	75	100	25	55	72	12	15	25	14	4	41	41	173	865
1991	240	475	2953	111	130	266	42	95	65	26	62	77	8	8	7	12	7	52	41	173	2486
1992	230	475	2965	109	145	272	38	65	55	28	67	82	7	11	11	7	14	59	41	173	2486
1993	242	452	2981	120	134	287	34	49	39	32	76	89	9	4	11	6	16	69	41	173	2486
1994	279	474	3042	136	137	331	40	49	57	38	86	100	8	9	9	16	20	59	41	173	2486
1996	345	532	3173	143	148	417	66	70	48	48	101	114	19	8	21	28	32	87	41	173	2486
1998	380	637	3270	147	166	418	89	115	68	51	115	129	30	44	62	22	24	107	41	173	2486
2000	468	736	3449	193	255	589	117	125	111	62	133	140	37	37	49	18	13	74	41	173	2486

¹ Two groups of NLSY79 respondents have been dropped from interviewing: (1) 1,079 members of the 1,280 military subsample were dropped after the 1984 survey and (2) the 1,643 members of the supplemental economically disadvantaged non-black/non-Hispanic subsample were dropped after the 1990 interview. Of the 1,643 supplemental economically disadvantaged, non-black/non-Hispanic subsample members dropped, 22 remain classified as deceased.

Table 2.6.4 Reasons for Noninterview by Sample Type: NLSY79 1980–2000

Key C: Cross-sectional, S: Supplemental, M: Military

Survey Year	Total Not Interviewed			Reason for Noninterview																	
				Refusal			Can't Locate			Deceased			Other			Difficult Cases			Dropped ¹		
	C	S	M	C	S	M	C	S	M	C	S	M	C	S	M	C	S	M	C	S	M
1980	238	220	87	153	91	9	60	101	56	4	5	0	21	23	22	0	0	0	0	0	0
1981	219	187	85	133	71	16	30	64	20	15	14	0	41	38	49	0	0	0	0	0	0
1982	235	259	69	86	73	18	56	123	30	24	19	1	7	25	18	62	19	2	0	0	0
1983	209	202	54	103	94	23	43	63	18	27	26	4	15	14	8	21	5	1	0	0	0
1984	297	255	65	204	138	32	54	73	24	30	33	4	9	11	5	0	0	0	0	0	0
1985	360	338	1094	180	146	5	51	94	7	36	43	0	10	14	2	83	41	1	0	0	1079
1986	478	456	1097	284	230	10	78	115	7	43	51	1	14	22	0	59	38	0	0	0	1079
1987	573	527	1101	286	217	5	118	165	10	51	56	3	28	39	1	90	50	3	0	0	1079
1988	598	518	1105	335	248	4	107	128	13	56	68	3	43	36	4	57	38	2	0	0	1079
1989	540	442	1099	316	202	7	90	93	5	60	78	3	19	25	2	55	44	3	0	0	1079
1990	613	540	1097	385	269	8	101	139	6	67	82	3	23	28	1	37	22	0	0	0	1079
1991	555	2014	1099	316	182	9	97	99	6	74	87	4	9	13	1	59	12	0	0	1621	1079
1992	558	2015	1097	323	196	7	82	70	6	80	93	4	12	16	1	61	19	0	0	1621	1079
1993	574	2002	1099	338	192	11	57	62	3	89	104	4	13	10	1	77	13	1	0	1621	1079
1994	654	2039	1102	398	197	9	78	59	9	103	117	4	12	14	1	63	31	1	0	1621	1079
1996	821	2124	1105	486	216	6	86	87	11	128	130	5	23	24	1	98	46	3	0	1621	1079
1998	952	2230	1105	490	233	8	117	146	9	151	139	5	75	60	1	119	31	3	0	1621	1079
2000	1162	2374	1117	689	333	15	162	180	11	170	158	7	58	61	4	83	21	1	0	1621	1079

¹ Two groups of NLSY79 respondents have been dropped from interviewing: (1) 1,079 members of the 1,280 military subsample were dropped after the 1984 survey and (2) the 1,643 members of the supplemental economically disadvantaged, non-black/non-Hispanic subsample were dropped after the 1990 interview. Of the 1,643 supplemental economically disadvantaged, non-black/non-Hispanic sample members dropped after 1990, 22 remain classified as deceased.

2.7 Sample Representativeness & Attrition

This section reviews the number of respondents by race, sex, and NLSY79 sample type who have continued to be interviewed during all surveys. It also takes a brief look at the racial composition of the cohort at the initial and latest survey points.

Table 2.7.1 shows the number of respondents, excluding dropped respondents, who were interviewed at all survey points. This table exhibits the high degree of NLSY79 retention. From 1979 to 2000 the survey has been administered 19 times; excluding dropped individuals, the average respondent has completed 17.4 interviews.

Table 2.7.1 Percentage of NLSY79 Respondents, Excluding Dropped Respondents, Who Answered Every Survey: 1979–2000

Year	Percent	Number	Year	Percent	Number
1979	100%	9964	1989	78.5%	7819
1980	96.0%	9571	1990	76.7%	7642
1981	94.3%	9395	1991	75.5%	7521
1982	92.7%	9234	1992	74.2%	7396
1983	91.6%	9125	1993	73.2%	7291
1984	89.7%	8942	1994	71.8%	7153
1985	87.5%	8721	1996	69.6%	6935
1986	85.0%	8472	1998	66.9%	6664
1987	82.3%	8203	2000	63.3%	6310
1988	79.9%	7957			

Table 2.7.2 compares the racial composition in 1979 versus 2000 of all sample members who have not been dropped. Additionally, it depicts the number of respondents in each racial/ethnic group who have attrited due to death.

Table 2.7.2 Cohort Characteristics by Race: 1979–2000

Race ¹	# of Interviewed Respondents		Racial Composition (percentage)		Retention (1998)	Number of Deaths
	1979	2000	1979	2000	as % of 1979	as of 2000
Hispanic	1961	1534	19.7	19.1	78.2	62
Black	3001	2438	30.1	30.3	81.2	133
Non-Black/Non-Hispanic	5002	4061	50.2	50.6	81.2	140

¹ See section on “Race, Ethnicity & Nationality” in this guide for details on race classifications.

Tables 2.7.3–2.7.5 show the distribution of the number of interviews completed by respondents, broken down by sex, race, and sample type. The “# who completed” column shows how many respondents completed **exactly** that number of surveys. These numbers refer to *any* surveys completed since the NLSY79 cohort started, not necessarily consecutive surveys completed or surveys completed in particular years. The cumulative percent column shows a cumulative total percent of those completing **at least** a given number of surveys rather than a percentage of those completing an **exact** number of surveys. Readers should note the attrition suggested in Table 2.7.3 greatly over represents the amount of lost information. NLSY79 surveys ask detailed questions about work history, education, training, marital status, and fertility since the date of the respondent’s last interview. These retrospective questions capture information lost due to missing interviews. Hence, a perfect response record is not needed for researchers to understand how the respondent’s life changes over time, unless he or she leaves the survey forever.

Table 2.7.3 Number of Interviews Respondents Completed out of 19 Surveys by Sex: NLSY79 1979–2000

Total			Males			Females		
# of Surveys ¹	# who Completed	Cumulative Percent	# of Surveys ¹	# who Completed	Cumulative Percent	# of Surveys ¹	# who Completed	Cumulative Percent
19	6310	100	19	2921	100	19	3389	100
18	1271	36.7	18	715	41.9	18	556	31.4
17	567	23.9	17	345	27.6	17	222	20.2
16	392	18.2	16	227	20.7	16	165	15.7
15	240	14.3	15	148	16.2	15	92	12.3
14	181	11.9	14	106	13.3	14	75	10.5
13	144	10.1	13	81	11.2	13	63	9.0
12	116	8.6	12	67	9.6	12	49	7.7
11	104	7.5	11	55	8.2	11	49	6.7
10	89	6.4	10	41	7.1	10	48	5.7
9	89	5.5	9	45	6.3	9	44	4.7
8	77	4.6	8	42	5.4	8	35	3.8
7	93	3.9	7	59	4.6	7	34	3.1
6	73	2.9	6	47	3.4	6	26	2.4
5	58	2.2	5	35	2.5	5	23	1.9
4	43	1.6	4	22	1.8	4	21	1.4
3	36	1.2	3	22	1.3	3	14	1.0
2	40	0.8	2	27	0.9	2	13	0.7
1	41	0.4	1	18	0.4	1	23	0.5
Total	9964	100	Total	5023	100	Total	4941	100

Note: Universe excludes the 1,079 members of the military subsample and the 1,643 members of the economically disadvantaged, non-black/non-Hispanic oversample dropped from interviewing; it includes the remaining 9,964 eligible members.

¹ Surveys completed in any year, not necessarily consecutive survey years.

Table 2.7.4 Number of Interviews Respondents Completed out of 19 Surveys by Race: NLSY79 1979–2000

Total Sample			Hispanic			Black			Non-Black/Non-Hispanic		
# of Surveys ¹	# who Completed	Cumul. Percent	# of Surveys ¹	# who Completed	Cumul. Percent	# of Surveys ¹	# who Completed	Cumul. Percent	# of Surveys ¹	# who Completed	Cumul. Percent
19	6310	100	19	1087	100	19	1880	100	19	3343	100
18	1271	36.7	18	307	44.6	18	426	37.4	18	538	33.2
17	567	23.9	17	151	28.9	17	180	23.2	17	236	22.4
16	392	18.2	16	95	21.2	16	127	17.2	16	170	17.7
15	240	14.3	15	56	16.4	15	71	12.9	15	113	14.3
14	181	11.9	14	46	13.5	14	45	10.6	14	90	12.0
13	144	10.1	13	23	11.2	13	39	9.1	13	82	10.2
12	116	8.6	12	24	10.0	12	30	7.8	12	62	8.6
11	104	7.5	11	31	8.8	11	32	6.8	11	41	7.4
10	89	6.4	10	14	7.2	10	35	5.7	10	40	6.5
9	89	5.5	9	21	6.5	9	24	4.5	9	44	5.7
8	77	4.6	8	24	5.4	8	18	3.7	8	35	4.9
7	93	3.9	7	18	4.2	7	19	3.1	7	56	4.2
6	73	2.9	6	17	3.3	6	19	2.5	6	37	3.0
5	58	2.2	5	11	2.4	5	13	1.9	5	34	2.3
4	43	1.6	4	11	1.8	4	14	1.4	4	18	1.6
3	36	1.2	3	7	1.3	3	7	1.0	3	22	1.3
2	40	0.8	2	7	0.9	2	13	0.7	2	20	0.8
1	41	0.4	1	11	0.6	1	9	0.3	1	21	0.4
Total	9964	100	Total	1961	100	Total	3001	100	Total	5002	100

Note: Universe excludes the 1,079 members of the military subsample and the 1,643 members of the economically disadvantaged, non-black/non-Hispanic oversample dropped from interviewing; it includes the remaining 9,964 eligible members.

¹ Surveys completed in any year, not necessarily consecutive survey years.

Table 2.7.5 Number of Interviews Respondents Completed out of 19 Surveys by Sample Type: NLSY79 1979–2000

Total Sample			Cross Sectional Sample			Supplemental Sample			Military Sample		
# of Surveys ¹	# who Completed	Cumul. Percent	# of Surveys ¹	# who Completed	Cumul. Percent	# of Surveys ¹	# who Completed	Cumul. Percent	# of Surveys ¹	# who Completed	Cumul. Percent
19	6310	100	19	4021	100	19	2190	100	19	99	100
18	1271	36.7	18	688	34.2	18	550	40.0	18	33	50.8
17	567	23.9	17	306	22.9	17	243	25.0	17	18	34.3
16	392	18.2	16	217	17.9	16	154	18.3	16	21	25.4
15	240	14.3	15	138	14.4	15	93	14.1	15	9	14.9
14	181	11.9	14	116	12.1	14	60	11.6	14	5	10.5
13	144	10.1	13	102	10.2	13	40	9.9	13	2	8.0
12	116	8.6	12	74	8.6	12	39	8.8	12	3	7.0
11	104	7.5	11	56	7.4	11	45	7.8	11	3	5.5
10	89	6.4	10	52	6.4	10	37	6.5	10	0	4.0
9	89	5.5	9	53	5.6	9	36	5.5	9	0	4.0
8	77	4.6	8	45	4.7	8	31	4.5	8	1	4.0
7	93	3.9	7	63	4.0	7	29	3.7	7	1	3.5
6	73	2.9	6	42	3.0	6	29	2.9	6	2	3.0
5	58	2.2	5	43	2.3	5	14	2.1	5	1	2.0
4	43	1.6	4	20	1.6	4	21	1.7	4	2	1.5
3	36	1.2	3	24	1.2	3	12	1.1	3	0	0.5
2	40	0.8	2	24	0.8	2	15	0.8	2	1	0.5
1	41	0.4	1	27	0.4	1	14	0.4	1	0	0.0
Total	9964	100	Total	6111	100	Total	3652	100	Total	201	100

Note: Universe excludes the 1,079 members of the military subsample and the 1,643 members of the economically disadvantaged, non-black/non-Hispanic oversample dropped from interviewing; it includes the remaining 9,964 eligible members.

¹ Surveys completed in any year, not necessarily consecutive survey years.

2.8 Sample Weights

In each survey year a set of sampling weights is constructed. These weights provide the researcher with an estimate of how many individuals in the United States each respondent's answers represent. Weighting decisions for the NLSY79 are guided by the following principles: (1) individual case weights are assigned for each year in such a way as to produce group population estimates when used in tabulations, and (2) the assignment of individual respondent weights involves at least three types of adjustment, with additional considerations necessary for weighting of NLSY79 Child data.

The interested user should consult the NLSY79 *Technical Sampling Report* (Frankel, Williams, and Spencer 1983) for a step-by-step description of the adjustment process. A cursory review of the process follows.

Adjustment One: The first weighting adjustment involves the reciprocal of the probability of selection at the first interview. Specifically, this probability of selection is a function of the probability of selection associated with the household in which the respondent was located, as well as the subsampling (if any) applied to individuals identified in screening.

Adjustment Two: This process adjusts for differential response (cooperation) rates in both the screening phase and subsequent interviews. Differential cooperation rates are computed (and adjusted) on the basis of geographic location and group membership, as well as within-group subclassification.

Adjustment Three: This weighting adjustment attempts to correct for certain types of random variation associated with sampling as well as sample "undercoverage." These ratio estimations are used to conform the sample to independently derived population totals.

Sampling Weight Readjustments: Sampling weights for the main survey are readjusted by NORC to account for noninterviews each survey year. The readjustments are necessitated by differential nonresponse and use base year sample parameters for their creation, employing a procedure similar to that described above. The only exception occurs in the final stage of post-stratification. Post-stratification weights in survey rounds two and above have been recomputed on the basis of completed cases in that year's sample rather than the completed cases in the base year sample.

Child weights are based on mother weights with an adjustment factor used to account for different interview rates for children in various age, race, and sex groups. These factors use counts of children known to exist as well as estimates of fertility for women who have attrited. However, child weights do not adjust for children who were eligible to complete an assessment but did not, even though they completed other parts of the interview. Baker et al. (1993) and the current *Child/Young Adult Data Users Guide* present a detailed discussion of child sampling weights.

Because of the complicated sample design, weighted descriptive statistics are recommended whenever inferences are drawn for the total population of youth that the sample represents. However, caution should be exercised when interpreting small changes in population statistics across years, particularly when using the child weights.

2.9 Practical Usage of Weights

The application of sampling weights varies depending on the type of analysis being performed. If tabulating sample characteristics for a single interview year in order to describe the population being represented (i.e., compute sample means, totals, or proportions), researchers should weight the observations using the weights provided. For example, to estimate the average hours worked in 1987 by persons born in 1957 through 1964, simply use the weighted average of hours worked, where weight is the 1987 sample weight. These weights are approximately correct when used in this way, with item nonresponse possibly generating small errors. Other applications for which users may wish to apply weighting, but for which the application of weights may not correspond to the intended result include:

Samples Generated by Dropping Observations with Item Nonresponses: Often users confine their analysis to subsamples for which respondents provided valid answers to certain questions. In this case, a weighted mean will not represent the entire population, but rather those persons in the population who would have given a valid response to the specified questions. Item nonresponse because of refusals, don't knows, or invalid skips is usually quite small, so the degree to which the weights are incorrect is probably quite small. In the event that item nonresponse constitutes only a small proportion of the data for variables under analysis, population estimates (i.e., weighted sample means, medians, and proportions) would be reasonably accurate. However, population estimates based on data items that have relatively high nonresponse rates, such as family income, may not necessarily be representative of the underlying population of the cohort under analysis. For more information on item nonresponse in the NLSY79, see Chapter 5 of this *User's Guide*.

Data from Multiple Waves: Because the weights are specific to a single wave of the study, and because respondents occasionally miss an interview but are contacted in a subsequent wave, a problem similar to item nonresponse arises when the data are used longitudinally. In addition, occasionally the weights for a respondent in different years may be quite dissimilar, leaving the user uncertain as to which weight is appropriate. In principle, if a user wished to apply weights to multiple wave data, weights would have to be recomputed based upon the persons for whom complete data are available. In practice, if the sample is limited to respondents interviewed in a terminal or end point year, the weight for that year can be used.

Regression Analysis: A common question is whether one should use the provided weights to perform weighted least squares when doing regression analysis. Such a course of action may not lead to correct

estimates. If particular groups follow significantly different regression specifications, the preferred method of analysis is to estimate a separate regression for each group or to use dummy (or indicator) variables to specify group membership.

Users interested in calculating the population average effect of, for example, education upon earnings, should simply compute the weighted average of the regression coefficients obtained for each group, using the sum of the weights for the persons in each group as the weights to be applied to the coefficients. While least squares is an estimator that is linear in the dependent variable, it is nonlinear in explanatory variables, and so weighting the observations will generate different results than taking the weighted average of the regression coefficients for the groups. The process of stratifying the sample into groups thought to have different regression coefficients and then testing for equality of coefficients across groups using an F-test is described in most statistics texts.

Users uncertain about the appropriate grouping should consult a statistician or other person knowledgeable about the data set before specifying the regression model. Note that if subgroups have different regression coefficients, a regression on a random sample of the population would be misspecified.

2.10 Design Effects

Because the samples are multi-stage, stratified random samples instead of simple random samples, respondents tend to come in geographic clusters and clusters of persons tend to be alike in a variety of ways for a variety of reasons. (For more information on the sampling and screening process, users are referred to sections 2.2 and 2.3 of this chapter.) For example, there may be cultural differences by locality or ecological differences in labor market conditions. Depending upon the degree of this homogeneity, the conventionally computed standard deviations for the variables, which assume a simple random sample, may be too small. However, by controlling the rate at which particular strata are sampled, multi-stage, stratified random samples can improve upon simple random samples. The ratio of the correct standard error to the standard error computed under the assumption of a simple random sample is known as the design effect. The technical sampling report for the NLSY79 (Frankel, Williams, and Spencer 1983) and its addendum (CHRR) provide design effects for the various strata.

A single design effect that can be broadly applied to regression analysis cannot be constructed. To illustrate the approximate size of design effects in regression analysis, a regression of rate of pay for the CPS job in 1979 was estimated using race, sex, marital status, and education as explanatory variables. Assuming each of the roughly 200 PSUs has the same number of respondents in the sample of 5,724 persons with observed wages, the design effect was calculated to be 1.52; that is, the true standard errors were larger than the naively computed standard errors by a factor of 1.52. When this exercise was repeated for rate of pay on the CPS job in 1986, the design effect had fallen to 1.37.

This reduction reflects the fact that mobility tends to mix the respondents more uniformly through the country, reducing the clustering of the sample. Many of the persons who started out in the same PSU will have moved to different areas and, hence, no longer share unobservable labor market conditions. These shared unobservable labor market conditions are likely responsible for the spatial correlation of the error terms which generate design effects. Thus, another advantage of longitudinal data is the lessening of design effects over time.

By examining the geocode data for the NLSY79, it is possible to control for some of the environmental factors generating design effects or, if desired, to compute design effects based upon county or metropolitan area clusters which continue to be present. To facilitate study of design effects, scrambled PSU codes from the 1979 survey are available to persons ordering NLSY79 geocode data.

Users interested in information on how to use the standard errors and design effects should refer to Appendix B, “Standard Errors and Design Effects,” at the end of this guide. The Technical Sampling Report and Technical Sampling Report Addendum also provide information on standard errors and design effects.

References

- Center for Human Resource Research. “Technical Sampling Report Addendum: Standard Errors and Deft Factors for Rounds IV through XIV.” Columbus, OH: CHRR, The Ohio State University, 1994.
- Frankel, M.R.; Williams, H.A.; and Spencer, B.D. *Technical Sampling Report, National Longitudinal Survey of Labor Force Behavior*. Chicago: NORC, University of Chicago, 1983.
- Baker, Paula C.; Mott, Frank L.; Keck, Canada K.; and Quinlan, Stephen V. *NLSY79 Child Handbook: A Guide to the 1986–1990 NLSY79 Child Data*. Columbus, OH: CHRR, The Ohio State University, 1993.
- NORC. *NLSY-National Longitudinal Survey of Labor Force Behavior Interviewer’s Manual-Household Screening*. Chicago: NORC, University of Chicago, 1978.
- Olsen, Randall J. “The Effects of Computer Assisted Interviewing on Data Quality.” Columbus, OH: CHRR, The Ohio State University, 1991.

